تخمین مهم‌ترین پارامتر ژئومکانیکی خاک‌های غیراشباع با به‌کارگیری روش‌های مختلف هوش مصنوعی*

مقاله پژوهشی

هادی فتاحی[1]　　فاطمه جیریایی[2]

چکیده *در پیش‌بینی رفتار سازه‌های ژئوتکنیکی که بر خاک‌های غیر اشباع احداث می‌شوند، دانستن خواص ژئومکانیکی به‌ویژه تنش مؤثر از اهمیت ویژه‌ای برخوردار است. تأثیر مکش در خاک های غیراشباع موجب تغییر در رفتار حجمی و برشی خاک می‌شود. از این رو در تعیین مقاومت برشی در خاک‌های غیراشباع، پارامتر تنش مؤثر (χ) به‌عنوان تابعی از مکش نقش اساسی دارد. تعیین این پارامتر نیازمند زمان و هزینه زیادی است که در آزمون‌های آزمایشگاهی صرف می‌شود. هدف از این تحقیق ارزیابی چند روش هوشمند برای ایجاد مدل‌هایی است که از طریق غیرمستقیم پارامتر χ را به‌طور دقیق تخمین بزنند. بدین منظور از ۱۲۰ داده (که از نتایج آزمایش‌های سه محوره، برشی، صفحه فشار و کاغذ فیلتر است) و روش‌های هوشمند جنگل تصادفی، ماشین بردار پشتیبان و نزدیک‌ترین همسایگی در نرم‌افزار هوشمند WEKA استفاده شده‌است. نتایج نشان می‌دهد که مدل‌های توسعه یافته توسط سه روش هوشمند جنگل تصادفی، ماشین بردار پشتیبان و نزدیک‌ترین همسایگی، از عملکرد و دقت بسیار خوبی برخوردار هستند. اما مدل جنگل تصادفی در ارزیابی بر روی داده‌های آزمون با $R^2=0.918$ و $RMSE=0.079$، بهتر از دو مدل دیگر است. به‌علاوه در این تحقیق آنالیز حساسیت به‌منظور تعیین اهمیت پارامترهای مؤثر بر روی پارامتر χ انجام شد که از بین پارامترهای ورودی در مدل‌سازی، مشخص شد که پارامتر محتوای آب حجمی ($\frac{\theta_r}{\theta_s}$) بیشترین تأثیر را بر روی پارامتر χ دارد.*

واژه‌های کلیدی تنش مؤثر، جنگل تصادفی، ماشین بردار پشتیبان، نزدیک‌ترین همسایگی، آنالیز حساسیت، نرم‌افزار WEKA.

مقدمه

امروزه مهندسان در اکثر پروژه‌های معدنی و عمرانی با خاک‌های غیراشباع سر و کار دارند. پروژه‌هایی مانند احداث جاده‌ها، فضاهای زیرزمینی و تونل‌ها، ساخت سدهای خاکی و سنگ‌ریزه‌ای و فرودگاه‌ها، خاک‌ریزهای شیروانی‌ها و... از مواردی هستند که در حین عملیات با خاک‌های غیراشباع برخورد می‌شود [1]. خاک غیراشباع (Unsaturated Soil) یک محیط متخلخل چندفازی است که از دانه‌های خاک و فاز مایع و گاز که معمولاً آب و هواست، تشکیل می‌شود. از این رو، وجود نیروی کشش سطحی در مرز تماس آب و هوا و یا به عبارتی وجود پارامتر مکش (Suction)، فشار فاز سیال را متفاوت از حالت خاک اشباع می‌کند و منجر به تغییر در رفتار خاک می‌شود [2]. باران، تبخیر یا بالا آمدن سطح آب زیرزمینی، از عوامل زیست‌محیطی هستند که منجر به خشک‌شدگی یا ترشدگی و به‌عبارتی اشباع یا غیراشباع بودن خاک می‌شوند. مدل‌سازی رفتاری خاک و طراحی‌ها بدون در نظر گرفتن شرایط خاک غیراشباع، گرچه ساده‌تر انجام می‌شود اما ضریب ایمنی را کاهش و هزینه‌ها را افزایش می‌دهد. اصل تنش مؤثر (Effective Stress) یک اصل اساسی در مکانیک خاک است که یک محیط متخلخل متشکل از چند فاز و چند تنشی را به یک محیط پیوسته تشکیل شده از یک فاز و یک تنش تبدیل می‌کند تا امکان استفاده از اصول مکانیک جامد فراهم شود. تعیین مقاومت برشی در خاک‌های غیراشباع در رابطه بیشاپ (Bishop)، علاوه‌بر تنش مؤثر وابسته به پارامتری به نام پارامتر تنش مؤثر (χ) است. این پارامتر، میزان مشارکت مکش در تنش مؤثر را مشخص می‌کند. مقدار این پارامتر در حالت اشباع یک و در حالت خشک صفر است [3]. رابطه بین پارامتر χ و پارامترهای مرتبط آن یک رابطه غیر خطی و پیچیده است و تعریف این رابطه به‌طور تجربی با خطای زیادی همراه می‌شود. با استفاده از رویکردهای پیشرفته‌ای مانند روش‌های هوشمند و سیستم‌های استنباطی می‌توان به

* تاریخ دریافت مقاله ۱۴۰۰/۹/۱۸ و تاریخ پذیرش آن ۱۴۰۱/۲/۲۶ می‌باشد.

(۱) نویسندهٔ مسئول، دانشیار، دانشکدهٔ مهندسی علوم زمین، دانشگاه صنعتی اراک.
(۲) دانشجو، دانشکدهٔ مهندسی علوم زمین، دانشگاه صنعتی اراک.
Email: h.fattahi@arakut.ac.ir

پیش‌بینی‌های دقیقی از χ دست یافت. پژوهش‌هایی که در سال‌های اخیر در این راستا انجام شده‌است به شرح زیر است.

لی و همکاران (۲۰۰۳) [4] یک رابطه غیرخطی برای تخمین مقاومت برشی در خاک غیر اشباع ارائه دادند که در آن برای پیش‌بینی پارامتر چسبندگی C_{max} از روش شبکه عصبی مصنوعی (Artificial Neural Network) استفاده کردند.

کایادلن [5] با هدف بررسی قابلیت شبکه‌های عصبی مصنوعی برای مدل‌سازی رابطه پیچیده بین پارامترهای اساسی خاک، مکش و پارامتر χ، پنج مدل شبکه عصبی مصنوعی را با پارامترهای ورودی متفاوت ایجاد کرد.

اجدری و همکاران [6] از مدل شبکه عصبی مصنوعی برای پیش‌بینی پارامتر تنش مؤثر استفاده کردند.

جوهری و همکاران [7] به توسعه یک مدل GEP (Gene Expression Programing) به‌عنوان یک مدل شبکه عصبی مصنوعی برای پیش بینی پارامتر χ پرداختند. آن‌ها نشان دادند، پارامتر مکش مؤثرترین پارامتر در پیش‌بینی پارامتر χ است.

کاننگو و همکاران [8] روش‌های شبکه عصبی مصنوعی و رگرسیون درختی CART (Classification and Regression Tree) را برای تخمین غیرمستقیم پارامترهای مقاومت برشی خاک غیر اشباع به‌کار بردند. آن‌ها برای ساخت مدل‌های مختلف از پارامترهای ورودی درصد تشکیل‌دهنده شن، ماسه، سیلت، خاک رس، چگالی خشک و شاخص پلاستیسیته استفاده کردند.

راهنما و همکاران [9] دو مدل سیستم استنتاج فازی FCM-FIS (Fuzzy C-means Clustering method- Fuzzy) و SC-FIS (Subtractive clustering) را به‌کار گرفتند و پارامتر χ را پیش‌بینی کردند. آن‌ها نشان دادند پارامترهای فشار محصورکننده (Confining Pressure) و محتوای آب حجمی (Volumetric Water Content) مؤثرترین پارامترها در روند پیش‌بینی هستند.

تین بوی و همکاران [10] یک روش پیشنهادی برای پیش‌بینی مقاومت برشی خاک ارائه دادند و در آن دو روش حداقل مربعات ماشین بردار پشتیبان LSSVM (Least Squares Support Vector Machine) و الگوریتم CSO (Cuckoo Search Optimization) را تلفیق کردند. آن‌ها برای ارزیابی مدل ارائه شده از نمونه‌های جمع‌آوری شده از یک بزرگراه در ویتنام استفاده کردند.

در این تحقیق برای پیش بینی پارامتر χ به‌واسطه پارامترهای مؤثر آن، از روش‌های هوش مصنوعی نظیر، جنگل تصادفی، ماشین بردار پشتیبان و نزدیکترین همسایگی، برای مدل‌سازی استفاده می‌شود. مدل‌سازی‌ها در نرم افزار هوشمند WEKA انجام می‌شود و در آن‌ها مجموعه داده‌های آزمایشگاهی جمع آوری شده از چندین مقاله به کار گرفته خواهد شد. پارامترهای ورودی در این مدل، پارامتر منحنی مشخصه آب–خاک (Soil-Water Characteristic Curve Fitting Parameter)، فشار محصور کننده، مکش و محتوای آب حجمی و متغیر خروجی، پارامتر χ است.

پارامتر تنش مؤثر در روابط تجربی

در پروژه‌هایی که با خاک‌های غیراشباع سر و کار دارند، متغیر حالت تنش، نقش تعیین‌کننده‌ای در مدل‌سازی رفتاری خاک دارد. در خاک‌های اشباع تنش مؤثر به‌عنوان متغیر حالت تنش در نظر گرفته می‌شود، اما در خاک‌های غیراشباع متغیر حالت تنش، مطابق رابطه (۱) از دو متغیر حالت تنش مجزا تشکیل می‌شود.

$$\sigma' = (\sigma - u_a) + \chi(u_a - u_w) \qquad (۱)$$

که در آن χ پارامتر تنش مؤثر، u_a فشار آب حفره‌ای، $(\sigma - u_a)$ تنش خالص نرمال، $(u_a - u_w)$ نماینده مکش ارائه شده توسط خلیلی و خباز می‌باشد [11]. آن‌ها χ را به‌عنوان تابعی از نسبت مکش طبق رابطه (۲) معرفی کردند.

$$\chi = \begin{cases} \left(\frac{u_a - u_w}{u_e}\right)^{-0n55} & for \quad u_a - u_w > u_e \\ 1 & for \quad u_a - u_w \le u_e \end{cases} \qquad (۲)$$

که در آن u_e مقدار ورودی هوا می‌باشد. این رابطه گذار از حالت اشباع، به حالت غیراشباع را نشان می‌دهد. تعیین پارامتر χ نیازمند زمان و هزینه زیادی است که در طی آزمون‌های آزمایشگاهی صرف می‌شود. طوری که مطالعات نشان می‌دهد این پارامتر بسیار غیرخطی بوده و روابط تجربی قادر به تخمین دقیق آن نخواهند بود [12]. بیشاپ رابطه (۳) را برای محاسبه مقاومت برشی خاک‌های غیراشباع ارائه داد [13].

$$\tau = c' + (\sigma - u_a)\tan\varphi' + \chi s \tan\varphi \qquad (۳)$$

که در آن $\acute{c}$ و $\acute{\varphi}$ چسبندگی و زاویه اصطکاک داخلی مؤثر هستند. به‌طور کلی χ تابعی از درجه اشباع‌شدگی (sr) فرض می‌شود. رگرسیون‌های برازش‌شده بر روی برخی از داده‌های تجربی، تناسب خوبی را بین این دو پارامتر نشان می‌دهند. در رابطه (٤) χ تابعی از (sr) تعریف شده‌است [14].

$$\chi = S_r^k = \left(\frac{\theta}{\theta_s}\right)^k \qquad (٤)$$

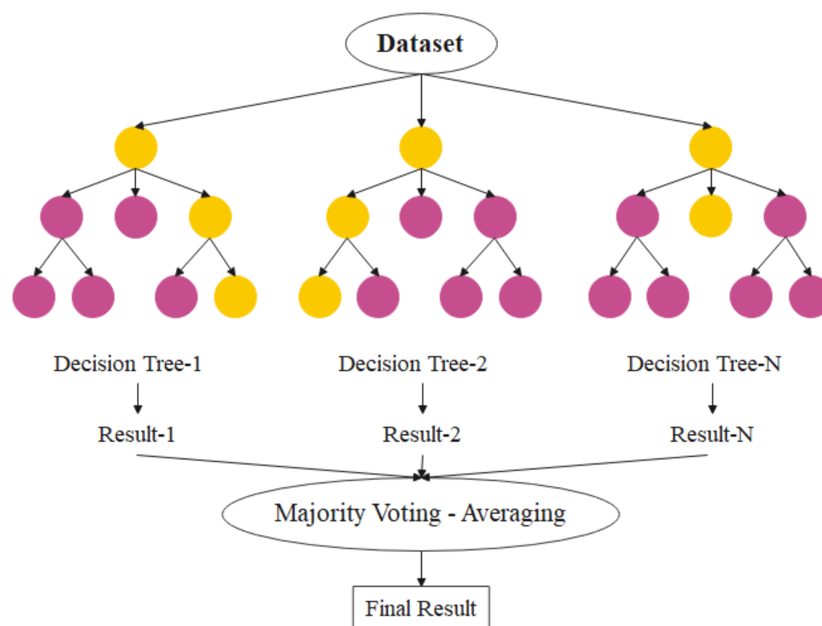
که در آن θ محتوای آب حجمی، θ_s محتوای آب اشباع و k یک پارامتر بهینه شده‌است که از طریق بهترین برازش بین داده‌های اندازه‌گیری‌شده و پیش‌بینی‌شده به‌دست می‌آید.

معرفی روش‌های هوشمند

در این بخش به توصیف برخی از تکنیک‌های داده‌کاوی که در این تحقیق بکارگرفته شده‌اند، پرداخته خواهد شد.

جنگل تصادفی RF (Random Forest)

جنگل تصادفی روشی برگرفته از روش درخت تصمیم است و برای اهداف طبقه‌بندی و رگرسیون داده‌های گسسته و پیوسته مناسب می‌باشد. روش درخت تصمیم نوسانات بالایی را در نتایج خود نشان می‌دهد، به همین دلیل کار جنگل تصادفی کاهش نوسانات و واریانس براوردها و در نتیجه بهبود نتایج درخت تصمیم است. جنگل تصادفی این کار را با ترکیب چند درخت تصمیم انجام می‌دهد، که در ساخت هر درخت تصمیم، تعدادی از متغیرها به‌طور تصادفی شرکت می‌کنند. مجموعه درخت‌ها با در نظر گرفتن n نمونه بوت استرپ از داده‌ها به‌عنوان داده‌های مشاهده‌ای مستقل ساخته می‌شوند. به این معنی که تعداد زیادی نمونه‌برداری همراه با جای‌گذاری انجام می‌شود. سپس بر روی هر نمونه یک درخت تصمیم تعریف می‌شود. اگر تعداد تمام متغیرهای مستقل M باشد، در هر شاخه از درخت به‌طور تصادفی m متغیر مستقل، برای تقسیم شدن انتخاب می‌شود. برای حالت رگرسیون نسبت $\frac{m}{M} = \frac{1}{3}$ و در حالت طبقه‌بندی $m = \sqrt{M}$ پیشنهاد می‌شود. سرانجام به تعداد درخت‌ها برای بردارهای ورودی، یک خروجی حاصل می‌شود و خروجی نهایی مدل از میانگین‌گیری این خروجی‌ها حاصل می‌شود [15]. شکل (۱) نمونه‌ای از جنگل تصادفی تشکیل شده از N درخت رویداد را نشان می‌دهد.



شکل ۱ نمایی از جنگل تصادفی متشکل از N درخت رویداد

ماشین بردار پشتیبان SVM (Support Vector Machine)

ماشین بردار پشتیبان یک الگوریتم نظارت‌شده یادگیری ماشین است که از نوع مدل طبقه‌بندی ماشین بردار پشتیبان برای کلاس‌بندی داده‌ها و از نوع مدل رگرسیون بردار پشتیبان، برای حل مسائل پیش‌بینی استفاده می‌شود. روش رگرسیون بردار پشتیبان از الگوریتم‌های آموزشی تشکیل شده که بر اساس ماشین بردار پشتیبان طبقه‌بندی‌کننده کار می‌کند. این روش با بهره‌گیری از تئوری یادگیری آماری و مینیمم کردن ریسک ساختاری به مدلی برای پیش‌بینی می‌رسد. روش SVR (Support Vector Regression) نخستین بار توسط وپنیک ارائه شد [16]. در این روش برای آموزش داده‌ها از یک نگاشت غیرخطی استفاده می‌شود و داده‌ها به فضایی با ابعاد بالاتر منتقل می‌شوند و سپس برای ساخت شکل رگرسیون از تابع خطای ε در رابطه (۵) استفاده می‌شود.

$$L\left(y.f\left(x.a\right)\right)=\left|y-f\left(x.a\right)\right|_{\varepsilon}= \\ =\begin{cases}0 & if\ \left|y-f\left(x.a\right)\right|\le\varepsilon \\ \left|y-f\left(x.a\right)\right|-\varepsilon & Otherwise\end{cases} \quad (۵)$$

رابطه (۵) نشان می‌دهد که خطاهایی با مقادیر کمتر از ε در نظر گرفته نمی‌شوند. رابطه غیرخطی پارامترهای ورودی و خروجی مطابق رابطه (٦) به شکل یک تابع بهینه نوشته می‌شود.

$$f(x) = w^T\varphi(x) + b \quad (٦)$$

که در آن x بردار متغیر ورودی، w بردار وزن و b بایاس است. این تابع به‌عنوان تابع زیان ε- غیرحساس (ε Insensitive-Loss Function) شناخته می‌شود و حل آن می‌تواند به شکل مسئله درجه دوم در رابطه (۷) بیان شود.

$$\begin{cases} Min\frac{1}{2}w^Tw + C\sum_{i=1}^{L}(\xi_i - \xi_i^*) \\ s.t.\ y_i - w\cdot\varphi(x_i) - b \le \varepsilon + \xi_i \\ w\cdot\varphi(x_i) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i \ge 0 \\ \xi_i^* \ge 0 \end{cases} \quad (۷)$$

که در آن C پارامتر جریمه و ξ متغیر کمکی است. با معرفی ضرایب لاگرانژ و کاربرد تابع کرنل به‌جای ضرب داخلی عملگرها، مسئله بهینه‌سازی فوق ساده‌تر شده و رابطه (٦) به شکل رابطه (۸) در می‌آید.

$$f(x) = \sum_{x_i\in SV}(a_i - a_i^*)K(x_i.x_j) + b \quad (۸)$$

که در آن a_i و a_i^* ضرایب لاگرانژ و $k(x_i.x_j)$ تابع کرنل است [17].

نزدیک‌ترین همسایگی (KNN)

روش نزدیک‌ترین همسایگی یک روش آماری ناپارامتری است که برای مسائل طبقه‌بندی و رگرسیون مورد استفاده قرار می‌گیرد. این روش مبتنی بر مفهوم تشابه بوده و پیش‌بینی‌ها بر اساس موقعیت‌های مشابهی که در گذشته اتفاق افتاده، انجام می‌شود. به عبارتی با مشاهده متغیرهای مستقل در زمان واقعی، مدل به جستجوی الگوهای مشابه شرایط فعلی در سری تاریخی می‌پردازد و گزینه‌های محتمل را انتخاب می‌کند [18]. رابطه ۹ بردار 1×m مقادیر پیش‌بینی‌کننده x_j در زمان t را نشان می‌دهد.

$$Pr_{jt} = \left(x_{jt}\right) \qquad j = 1 \dots m \quad (۹)$$

هم‌چنین ماتریس n×m مقادیر متغیرهای پیش‌بینی‌کننده x_j در سری زمانی تاریخی به‌صورت رابطه (۱۰) است.

$$Pr_{j.(t-i)} = x_{j.(t-i)} \quad (۱۰)$$

$$j = 1 \dots m\ .\ i = 1 \dots n$$

با استفاده از تابع فاصله‌سنج Dist، در رابطه (۱۱)، فاصله بین بردار $Pr_{j,t}$ با سطرهای ماتریس $Pr_{j,(t-i)}$ محاسبه می‌شود.

$$Dist(t-i) = f(w_j.x_{j.(t-i)}.x_{jt}) \quad (۱۱)$$

که در آن j اندیس مربوط به متغیرهای پیش‌بینی‌کننده و i اندیس مربوط به گام زمانی در سری تاریخی است. w_j نیز وزن‌های متغیرهای پیش‌بینی‌کننده می‌باشد. در ادامه می‌بایست به‌منظور بهبود عملکرد مدل، از روش‌های مناسبی برای تخمین بهترین همسایه‌ها، توسعه توابع انتقال اطلاعات و توابع فاصله سنجی استفاده کرد [19].

معرفی اجمالی نرم‌افزار هوشمند WEKA

تا به امروز نرم‌افزارهای فراوانی در حوزه داده‌کاوی عرضه

شده‌اند که هر یک متناسب با نوع و فرمت داده‌هایی که پردازش می‌کنند، شامل الگوریتم‌های مختلفی می‌شوند. نرم‌افزار WEKA از جمله ابزارهای قدرتمند داده‌کاوی است و شامل الگوریتم‌های روز یادگیری ماشین و ابزارهای پیش‌پردازش داده‌ها می‌باشد. در این نرم‌افزار می‌توان روش‌های مختلف داده‌کاوی را بر روی داده‌ها آزمایش کرد و با ایجاد چند میز کار، نتایج مدل‌سازی برای روش‌های مختلف یا داده‌های متفاوت را به‌طور هم‌زمان مقایسه کرد. مجموعه روش‌ها برای مسائلی مانند رگرسیون، طبقه‌بندی، خوشه‌بندی، کاوش قواعد انجمنی و انتخاب ویژگی، در این نرم‌افزار گرد آمده‌اند. همه روش‌ها، ورودی‌های خود را با فرمت ARFF دریافت می‌کنند که می‌توان آن را به‌وسیله یک درخواست از پایگاه داده تولید کرد. به‌علاوه کتابی جامع و مرتبط با این نرم افزار [20] برای آشنایی بیشتر معرفی می‌شود.

توصیف پایگاه داده برای مدل‌سازی با روش‌های هوشمند

پایگاه داده شرکت داده شده در آنالیزها، شامل ١٢٠ داده از نتایج آزمایش‌های سه‌محوره، برشی، صفحه فشار و کاغذ فیلتر است که از چندین مقاله جمع‌آوری شده‌است [21-29]. این داده‌ها شامل ٧ متغیر شامل: مکش (s)، فشار حبابی (h_b)، فشار خالص محصور کننده (p)، محتوای آب باقی‌مانده (θ_r)، محتوای آب حجمی اشباع شده (θ_s)، پارامتر منحنی مشخصه آب–خاک (λ) و پارامتر تنش مؤثر (χ) می‌باشند. به‌منظور پردازش اولیه، پارامترهای s، h_b، p، θ_r و θ_s به‌صورت بی‌بعد وارد مدل‌سازی می‌شوند بدین صورت که از نسبت $\frac{P}{P_0}$ پارامتر فشار محصور کننده بدون بعد حاصل می‌شود (که در آن پارامتر P_0=101.325KPa است) و پارامتر مکش نیز توسط نسبت $\frac{s}{h_b}$ بدون بعد می‌شود.

هم‌چنین از نسبت $\frac{\theta_r}{\theta_s}$ پارامتر محتوای آب حجمی بدون بعد به‌دست می‌آید. بدین صورت در مدل‌سازی‌ها پارامترهای ورودی شامل $\frac{P}{P_0}$، $\frac{s}{h_b}$، $\frac{\theta_r}{\theta_s}$ و λ و پارامتر خروجی χ می‌شوند. جدول (١) مشخصات آماری داده‌های مربوط به این پارامترها را نشان می‌دهد. از این تعداد داده نرمالایز شده، ٨٠ درصد به‌عنوان داده‌های آموزش و ٢٠ درصد، به‌عنوان داده‌های آزمون جهت ارزیابی مدل‌ها به‌کار گرفته می‌شوند.

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(Y_{mea} - Y_{pre})^2}{\sum_{k=1}^{n} Y_{mea}^2 - \frac{\sum_{k=1}^{n} Y_{pre}^2}{n}} \quad (١٢)$$

$$MAE = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{Y_{mea} - Y_{pre}}{Y_{mea}}\right| \quad (١٣)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(Y_{mea} - Y_{pre})^2} \quad (١٤)$$

مدل‌سازی و تحلیل نتایج

به‌منظور ارزیابی عملکرد و دقت مدل‌های هوشمند چندین شاخص و معیار شامل ضریب هم‌بستگی مربع (R^2)، میانگین درصد خطای مطلق (MAE) و جذر میانگین خطای مربع (RMSE) معرفی شده‌اند. در تمامی این معادلات (معادلات ١٢ تا ١٤) Y_{mea} و Y_{pre} به‌ترتیب مقادیر اندازه‌گیری‌شده و پیش‌بینی‌شده هستند و n تعداد نمونه‌ها می‌باشند.

جدول ١ مشخصات آماری پارامترهای ورودی و خروجی مدل‌ها

پارامترها	مینیمم	ماکزیمم	میانگین	انحراف معیار
فشار محصورکننده ($\frac{P}{P_0}$)	٠	٣/٩٤٨	١/٢٤٣	١/٠٣٤
مکش ($\frac{S}{h_b}$)	٠	٣٠٠	٢٤/٢٦٢	٤٦/٤١٥
محتوای آب حجمی ($\frac{\theta_r}{\theta_s}$)	٠	٠/٧١٤	٠/٢٤٩	٠/٢٣١
منحنی مشخصه آب– خاک (λ)	٠/١٩	١١/٨٢	١/٢٩٤	٢/٤١٩
پارامتر تنش مؤثر (χ)	٠/٠٩١	١	٠/٦٨٩	٠/٢٧٤

مطابق با روابط فوق، هر قدر مقادیر MAE و RMSE به صفر نزدیک‌تر و مقدار R^2 به‌سمت عدد ۱ نزدیک‌تر باشد، مدل پیش‌بینی‌شده به مقدار واقعی نزدیک‌تر شده و به‌عنوان یک مدل ایدئال پذیرفته می‌شود.

در ادامه، پس از آماده‌سازی داده‌ها نوبت به ساخت مدل‌هایی برای پیش‌بینی/تخمین پارامتر χ با استفاده از روش‌های هوشمند در نرم‌افزار WEKA می‌رسد. از جمله روش‌های هوشمند موجود در نرم‌افزار، سه روش هوشمند جنگل تصادفی، ماشین بردار پشتیبان و نزدیک‌ترین همسایگی که دارای عملکرد و دقت بهتری بوده‌اند، برای ساخت مدل‌های مدنظر مورد استفاده قرار گرفته‌اند. همان‌طور که قبلاً اشاره شد، برای ارزیابی عملکردها مدل‌ها از سه شاخص آماری استفاده می‌شود که در جدول (۲) مقادیر R^2، RMSE و MAE مربوط به هر یک از مدل‌ها، برای مجموعه داده‌های آموزش و آزمون به‌طور مجزا ارائه شده‌است. همان‌طور که مشخص است، ضریب R^2 برای هر سه مدل هوشمند در بخش آموزش داده‌ها، دارای مقادیر بالا و قابل توجهی است. البته مقادیر همین ضریب در بخش آزمون مدل‌های هوشمند، شرایط متفاوتی دارد. می‌توان گفت مدل جنگل تصادفی، در این بخش از شاخص‌های آماری مطلوبی برخوردار است.

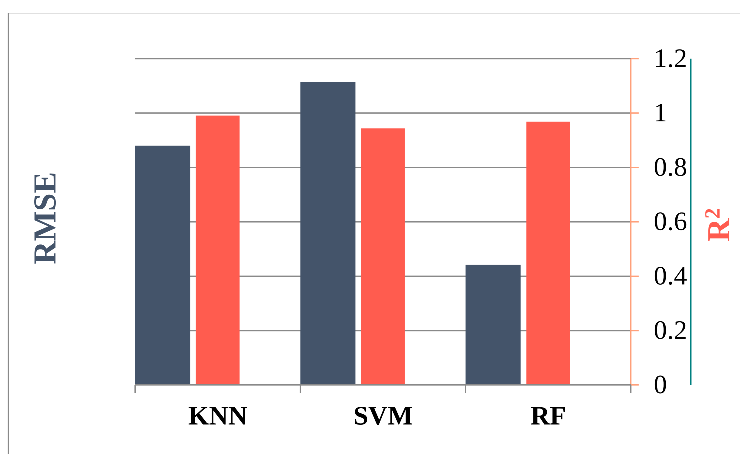
هم‌چنین نمودارهای نشان داده‌شده در شکل (۲) و (۳) به‌صورت گرافیکی مقادیر R^2 و RMSE سه مدل هوشمند جنگل تصادفی (RF)، ماشین بردار پشتیبان (SVM) و نزدیک‌ترین همسایگی (KNN) را برای مجموعه داده آموزش و آزمون مورد ارزیابی قرار می‌دهد. در بخش داده‌های آموزش مدل نزدیک‌ترین همسایگی به‌طور نسبی دارای مقدار R^2 بالا و RMSE پایین‌تری می‌باشد و از این نظر مدل جنگل تصادفی پس از مدل نزدیک‌ترین همسایگی قرار می‌گیرد اما با نظر به نمودار بخش داده‌های آزمون، مدل جنگل تصادفی دارای مقدار R^2 بالاتر و RMSE پایین‌تری است. بنابراین مدل جنگل تصادفی ضمن عملکرد و دقت خوب در داده‌های آموزش، همان عملکرد را در داده‌های آزمون نیز حفظ کرده‌است.

هوشمند برای داده‌های آزمون

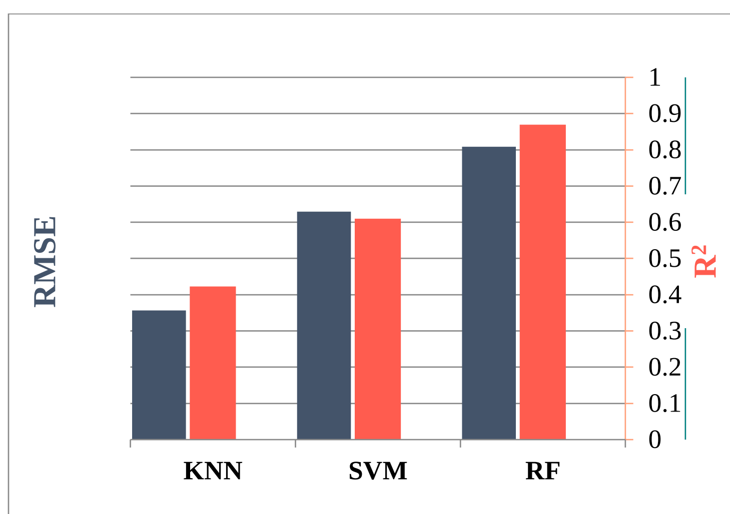
به‌علاوه در شکل‌های (٤) تا (٦) نیز می‌توان عملکرد مدل‌های هوشمند در داده‌های آموزش و آزمون را در کنار هم مشاهده کرد. در این نمودارها نقاط مقادیر اندازه‌گیری‌شده و پیش‌بینی‌شده پارامتر χ رسم شده‌است. در شکل مربوط به مدل جنگل تصادفی در هر دو نمودار، نقاط نزدیک به خط، قطری و دارای برازش مناسبی هستند اما در مدل نزدیک‌ترین همسایگی، نمودار داده‌های آموزش بسیار منظم و روی خط قطری اما در داده‌های آزمون بسیار پراکنده هستند. بنابراین مدل جنگل تصادفی در هر دو مجموعه دارای عملکرد بهتر و دقت بالاتری است و در مجموع مدل کارآمدی می‌باشد.

جدول ۲ مقدار شاخص‌های ارزیابی مدل‌ها در داده‌های آموزش و آزمون

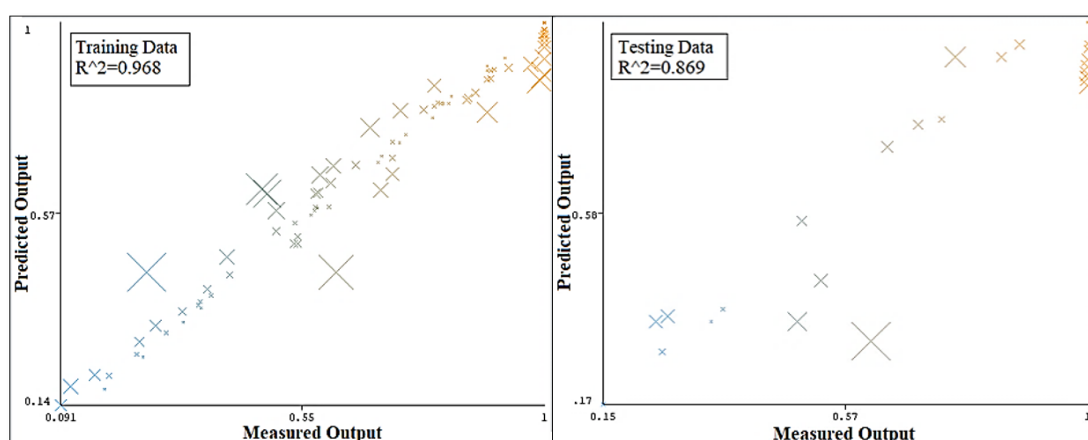
مدل‌ها		R^2	RMSE	MAE
جنگل تصادفی	برای داده‌های آموزش	۰/۹٦۷۹	۰/۰۵۱۳	۰/۰۳٤۳
	برای داده‌های آزمون	۰/۸٦۹۲	۰/۱۰٦۹	۰/۰۸٦٦
ماشین بردار پشتیبان	برای داده‌های آموزش	۰/۹٤۳۰	۰/۰٦۵	۰/۰۱۰٦
	برای داده‌های آزمون	۰/٦۱۰۱	۰/۱۸۸۹	۰/۱۵۷۸
نزدیک‌ترین همسایگی	برای داده‌های آموزش	۰/۹۹۰۸	۰/۰۲۵۸	۰/۰۰۳۷
	برای داده‌های آزمون	۰/٤۲۳۰	۰/۲٤۲۵	۰/۱۹۱



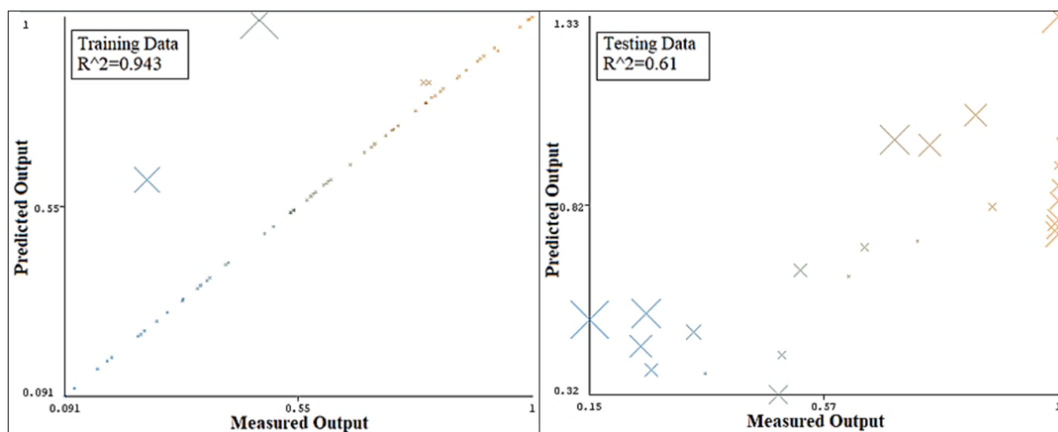
شکل ۲ نمودار ستونی مقادیر R^2 و RMSE سه مدل هوشمند برای داده‌های آموزش



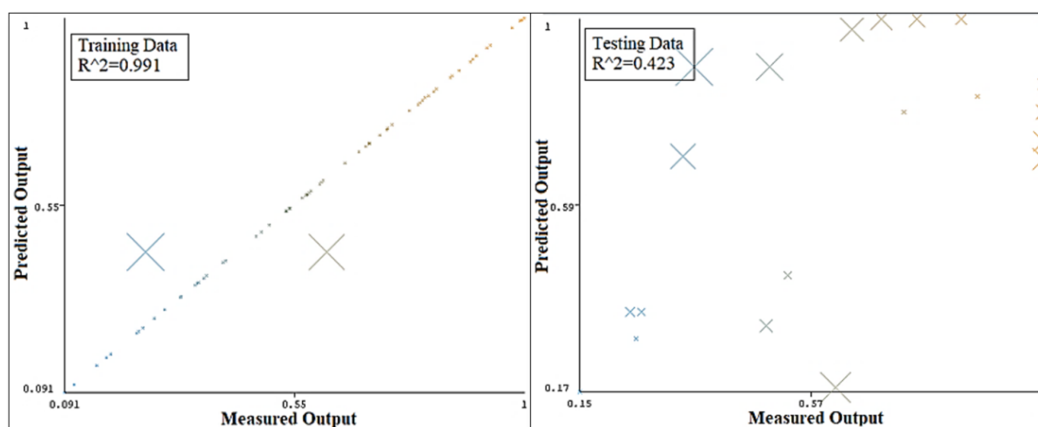
شکل ۳ نمودار ستونی مقادیر R^2 و RMSE سه مدل



شکل ٤ نمودار مقادیر اندازه‌گیری‌شده χ و مقادیر پیش بینی‌شده آن در مدل جنگل تصادفی در نرم‌افزار WEKA



شکل ٥ نمودار مقادیر اندازه‌گیری‌شده χ و مقادیر پیش‌بینی‌شده آن در مدل ماشین بردار پشتیبان در نرم‌افزار WEKA



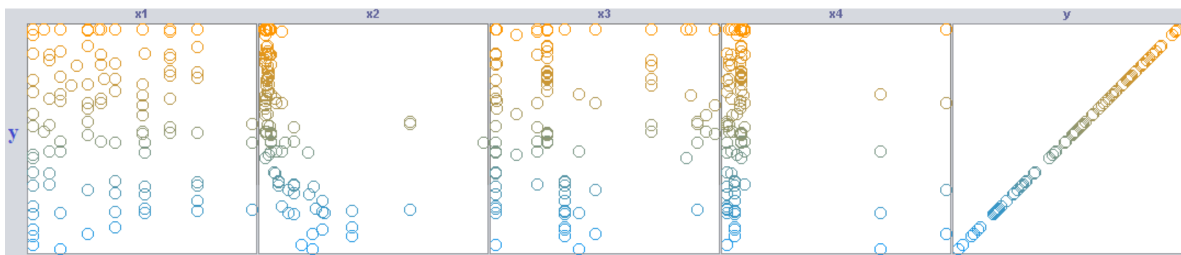
شکل ٦ نمودار مقادیر اندازه‌گیری‌شده χ و مقادیر پیش‌بینی‌شده آن در مدل نزدیک‌ترین همسایگی در نرم‌افزار WEKA

مدل‌سازی‌های انجام‌شده در قسمت‌های بالا، پارامترهای ورودی، هر یک با تاثیر متفاوتی در پیش‌بینی پارامتر χ مؤثر بوده‌اند. به‌منظور تشخیص تأثیرگذارترین پارامترها در تخمین، آنالیز حساسیتی در نرم‌افزار WEKA انجام شده‌است. ارزش و تأثیرگذاری هر پارامتر از طریق اندازه‌گیری هم‌بستگی پارامتر مورد نظر با پارامتر هدف تعیین می‌شود. در جدول (٤) با توجه به مقدار شایستگی متوسط، رتبه هر یک از پارامترها از لحاظ اهمیت مشخص شده‌است، که مطابق آن، $\frac{\theta_r}{\theta_s}$ مؤثرترین پارامتر و λ بی‌ارزش‌ترین پارامتر گزارش شده‌است.

هم‌چنین با رسم نمودار پراکندگی دو به دوی پارامترها می‌توان از چگونگی رابطه و رفتار آن‌ها نسبت‌به یکدیگر مطلع شد. شکل (٧) نمودار پراکندگی همه پارامترها را نسبت‌به پارامتر هدف یعنی پارامتر تنش مؤثر نشان می‌دهد. به این صورت که x_1، x_2، x_3 و x_4 به‌ترتیب بیانگر پارامترهای $\frac{P}{P_0}$، $\frac{s}{h_b}$، $\frac{\theta_r}{\theta_s}$ و λ و y بیانگر پارامتر χ است.

جدول ٣ آنالیز حساسیت برای تأثیر پارامترهای ورودی مدل‌های هوشمند در نرم‌افزار WEKA

پارامترها	رتبه	مقدار شایستگی متوسط
$\frac{\theta_r}{\theta_s}$	١	٠/١٥٧
$\frac{P}{P_0}$	٢	−٠/٠٥٦
$\frac{S}{h_b}$	٣	−٠/١٦
λ	٤	−٠/٥٢٥



شکل ۷ نمودار پراکندگی پارامترها نسبت به پارامتر خروجی در نرم‌افزار WEKA

نتیجه‌گیری

با توجه به این‌که درصد قابل توجهی از سطح زمین را خاک‌های غیراشباع دربرگرفته‌اند، بدیهی است در حین عملیات در اکثر پروژه‌های معدنی و عمرانی با این خاک‌ها برخورد می‌شود. تعیین مقاومت برشی در خاک‌های غیراشباع در رابطه بیشاپ وابسته به پارامتری به نام پارامتر χ است، که تابعی از مکش می‌باشد. در این تحقیق مدل‌هایی برای پیش‌بینی پارامتر χ ایجاد و ارزیابی شد. این مدل‌ها به‌کمک سه روش هوشمند جنگل تصادفی، ماشین بردار پشتیبان و نزدیک‌ترین همسایگی در قالب نرم‌افزار WEKA انجام شد و در آن از ۱۲۰ نمونه آزمایشگاهی جمع‌آوری‌شده از چندین مقاله استفاده شد. پارامترهای ورودی شامل پارامتر منحنی مشخصه آب- خاک (soil-water characteristic curve fitting parameter)، فشار محصور کننده، مکش و محتوای آب حجمی و پارامتر خروجی، پارامتر تنش مؤثر است. نتایج نشان داد:

- مدل جنگل تصادفی هم در مجموعه داده‌های آموزش (R^2=0.968 و RMSE=0.051) و هم در داده‌های آزمون (R^2=0.869 و RMSE=0.107)، با دقت بالا قابلیت انجام پیش‌بینی دقیق پارامتر χ را دارد.
- مدل‌های ماشین بردار پشتیبان و نزدیک‌ترین همسایگی نیز در خصوص مجموعه داده‌های آموزش، عملکرد و دقت خوبی دارند.
- آنالیز حساسیت انجام‌شده بر روی پارامترها نشان داد، پارامتر محتوای آب حجمی اهمیت بالایی در پیش‌بینی پارامتر χ دارد و در مقابل، پارامتر منحنی مشخصه آب-خاک، نسبت‌به سایر پارامترها اهمیت کمتری دارد.
- به‌طور کل می‌توان گفت نرم‌افزار قدرتمند و کاربردی WEKA می‌تواند به‌عنوان یک ابزار قوی در علم داده‌کاوی و علوم مهندسی زمین در تخمین پارامترهای ژئومکانیکی به‌کار گرفته شود.

مراجع

1. Fredlund, D.G., "Unsaturated soil mechanics in engineering practice". *Journal of geotechnical geoenvironmental engineering*, Vol. 132, No. 3, pp. 286-321, (2006).
2. Aitchison, G.D., Russam, K., and Richards, B., "Engineering concepts of moisture equilibria and moisture changes in soils". *Ministry of Transport*, Vol., (1966).
3. Sheng, D., "Constitutive modelling of unsaturated soils: Discussion of fundamental principles". *Unsaturated soils*, Vol. 1, No. pp. 91-112, (2011).
4. Lee, S., Lee, S.R., and Kim ,Y., "An approach to estimate unsaturated shear strength using artificial neural network and hyperbolic formulation". *Computers Geotechnics*, Vol. 30, No. 6, pp. 489-503, (2003).
5. Kayadelen, C., "Estimation of effective stress parameter of unsaturated soils by using artificial neural networks". *International journal for numerical analytical methods in geomechanics*, Vol. 32, No. 9, pp. 1087-1106, (2008).

6. Ajdari, M., Habibagahi, G., and Ghahramani, A., "Predicting effective stress parameter of unsaturated soils using neural networks". *Computers and Geotechnics*, Vol. 40, No. pp. 89-96, (2012).

7. Johari, A., Habibagahi, G., and Nakhaee, M., "Prediction of unsaturated soils effective stress parameter using gene expression programming". *Scientia Iranica* ,Vol. 20, No. 5, pp. 1433-1444, (2013).

8. Kanungo, D., Sharma, S., and Pain, A., "Artificial Neural Network (ANN) and Regression Tree (CART) applications for the indirect estimation of unsaturated soil shear strength parameters". *Frontiers of earth science,* Vol. 8, No. 3, pp. 439-456, (2014).

9. Rahnema, H., Hashemi Jokar, M., and Khabbaz, H., "Predicting the effective stress parameter of unsaturated soils using adaptive neuro-fuzzy inference system". *Scientia Iranica*, Vol. 26, No. 6, pp. 3140-3158, (2019).

10. Tien Bui, D., Hoang, N.-D., and Nhu, V.-H., "A swarm intelligence-based machine learning approach for predicting soil shear strength for road construction: a case study at Trung Luong National Expressway Project (Vietnam)". *Engineering with Computers*, Vol. 35, No. 3, pp. 955-965, (2019).

11. Khalili, N. and Khabbaz, M., "A unique relationship for χ for the determination of the shear strength of unsaturated soils". *Geotechnique*, Vol. 48, No. 5, pp. 681-687, (1998).

12. Cui, Y.J., et al., "Unsaturated soil mechanics". *International Society of Soil Mechanics and Geotechnical Engineering*, Vol., (2004).

13. Bishop, A.W., "The principle of effective stress". *Teknisk ukeblad*, Vol. 39, No. pp. 859-863, (1959).

14. Vanapalli, S., et al., "Model for the prediction of shear strength with respect to soil suction". *Canadian geotechnical journal*, Vol. 33, No. 3, pp. 379-392, (1996).

15. vladimir, et al., "A classification and regression tool for compound classification and qsar modeling svetnik". *Journal of Chemical Information Computer Sciences*, Vol. 43, No. 6, pp. 1947-1958, (2003).

16. Cortes, C. and Vapnik, V., "Support vector machine". *Machine learning*, Vol. 20, No. 3, pp. 273-297, (1995).

17. Zhang, D., et al., "Parameter optimization for support vector regression based on genetic algorithm with simplex crossover operator". *Journal of Information Computational Science*, Vol. 8, No. 6, pp. 911-920, (2011).

18. Shahrabi, J. and Hejazi, T., *Data mining 2, Tehran, Industrial university of amirkabir*. 2013, Jahad daneshgahi Press.(In Persian).

19. Azmi, M. and Araghinejad, S., "Development of K-Nearest Neighbour Regression Method in Forecasting River Stream Flow". *Journal of Water and Wastewater*, Vol. 23, No. 2, pp. 108-119, (2012).

20. Witten, I.H., et al. Practical machine learning tools and techniques. in DATA MINING. 2005.

21. Thu, T.M., Rahardjo, H., and Leong, E.-C., Effects of hysteresis on shear strength envelopes from constant water content and consolidated drained triaxial tests, in Unsaturated soils 2 .٢٠٠٦p. 1212-1222.

22. Russell, A.R. and Khalili, N., "A bounding surface plasticity model for sands exhibiting particle crushing". *Canadian Geotechnical Journal*, Vol. 41, No. 6, pp. 1179-1192, (2004).

23. Russell, A. and Khalili, N., "A unified bounding surface plasticity model for unsaturated soils". *International Journal for Numerical Analytical Methods in Geomechanics*, Vol. 30, No. 3, pp. 181-212, (2006).

24. Rampino, C., Mancuso, C., and Vinale, F., "Experimental behaviour and modelling of an unsaturated compacted soil". *Canadian Geotechnical Journal*, Vol. 37, No. 4, pp. 748-763, (2000).

25. Rahardjo, H., Heng, O.B., and Choon, L.E., "Shear strength of a compacted residual soil from consolidated drained and constant water content triaxial tests". *Canadian Geotechnical Journal*, Vol. 41, No. 3, pp. 421-436, (2004).

26. Miao, L., Liu, S., and Lai, Y., "Research of soil–water characteristics and shear strength features of Nanyang expansive soil". *Engineering Geology*, Vol. 65, No. 4, pp. 261-267, (2002).

27. Lee, I.-M., Sung, S.-G., and Cho, G.-C., "Effect of stress state on the unsaturated shear strength of a weathered granite". *Canadian Geotechnical Journal*, Vol. 42, No. 2, pp. 624-631, (2005).

28. Khalili, N., Geiser, F., and Blight, G., "Effective stress in unsaturated soils: Review with new evidence". *International journal of Geomechanics*, Vol. 4, No. 2, pp. 115-126, (2004).

29. Bishop, A.W. and Blight, G., "Some aspects of effective stress in saturated and partly saturated soils". *Geotechnique*, Vol ,١٣ .No. 3, pp. 177-197, (1963).